

**Sequence Information:****SEQ ID NO: 1 (INSP005A nucleotide sequence exon 1)**

5 1 ATGGGTGGTA GTGGTGTGCGT GGAGGTCCCC TTCCCTGCTCT CCAGCAAGTA  
51 CG

**SEQ ID NO: 2 (INSP005A protein sequence exon 1)**

10 1 MGGSGVVEVP FLLSSKYD

**SEQ ID NO: 3 (INSP005A nucleotide sequence exon 2)**

15 1 ATGAGCCCAG CCGCCAGGTC ATCCTGGAGG CTCTTGCGGA GTTTGAACGT  
51 TCCACGTGCA TCAGGTTTGT CACCTATCAG GACCAGAGAG ACTTCATTTTC  
101 CATCATCCCC ATGTATGG

**SEQ ID NO: 4 (INSP005A protein sequence exon 2)**

20 1 EPSRQVILEA LAEFERSTCI RFVTYQDQRD FISIIPMYG

**SEQ ID NO: 5 (INSP005A nucleotide sequence exon 3)**

25 1 GTGCTTCTCG AGTGTGGGGC GCAGTGGAGG GATGCAGGTG GTCTCCCTGG  
51 CGCCCACGTG TCTCCAGAAG GGCCGGGGCA TTGTCCCTCA TGAGCTCATG  
101 CATGTGCTGG GCTTCTGGCA CGAGCACACG CGGGCCGACC GGGACCGCTA  
151 TATCCGTGTC AACTGGAACG AGATCCTGCC AG

**30 SEQ ID NO: 6 (INSP005A protein sequence exon 3)**

1 CFSSVGRSGG MQVVS LAPT C LQKGRGIVLH ELMHVLGF WH EHTRADRDY  
51 IRVNWNEILP G

**35 SEQ ID NO: 7 (INSP005A nucleotide sequence exon 4)**

1 GCTTTGAAAT CAACTTCATC AAGTCTCAGA GCAGCAACAT GCTGACGCC  
51 TATGACTACT CCTCTGTGAT GCACTATGGG AG

**40 SEQ ID NO: 8 (INSP005A protein sequence exon 4)**

1 FEINFIKSQS SNMLTPYDYS SVMHYGR

**SEQ ID NO: 9 (INSP005A nucleotide sequence exon 5)**

45 1 GCTCGCCTTC AGCCGGCGTG GGCTGCCAC CATCACACCA CTTTGGGCC  
51 CCAGTGTCCA CATCGGCCAG CGATGGAACC TGAGTGCCTC GGACATCACC  
101 CGGGTCCTCA AACTCTACGG CTGCAGCCA AGTGGCCCCA GGCCCCGTGG  
50 151 GAGAG

**SEQ ID NO: 10 (INSP005A protein sequence exon 5)**

1 LAFSRRGLPT ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG

5 51 RG

**SEQ ID NO: 11 (INSP005A nucleotide sequence exon 6)**

1 GGTCCCATG C CCACAGCACT GGTAGGAGCC CCGCCCCGGC CTCCCTATCT

10 51 CTGCAGCGGC TTTTGGAGGC ACTGTGGCG GAATCCAGGA GCCCCGACCC

101 CAGTGGTTCC AGTGCAGGAG GCCAGCCC GT TCCTGCAGGG CCTGGGGAGA

151 GCCCACATGG GTGGGAGTCC CCTGCCCTGA AAAAGCTCAG TGCAGAGGCC

15 201 TCGGCAAGGC AGCCTCAGAC CCTAGCTTCC TCCCCAAGAT CAAGGCCTGG

251 AGCAGGTGCC CCCGGTGTG CTCAGGAGCA GTCCTGGCTG GCCGGAGTGT

20 301 CCACCAAGCC CACAGTCCC TCTTCAGAAG CAGGAATCCA GCCAGTCCCT

351 GTCCAGGGAA GCCCAGCTCT GCCAGGGGGC TGTGTACCTA GAAATCATTT

401 CAAGGGGATG TCCGAAGAT

**25 SEQ ID NO: 12 (INSP005A protein sequence exon 6)**

1 SHAHSTGRSP APASLSLQRL LEALSAESRS PDPSGSSAGG QPVPAGPGES

51 PHGWESPALK KLSAEASARQ PQTLOSSPRS RPGAGAPGVA QEWSLAGVS

30 101 TKPTVPSSEA GIQPVPVQGS PALPGCCVPR NHFKGMSED

**SEQ ID NO: 13 (INSP005A full nucleotide sequence)**

1 ATGGGTGGTA GTGGTGTGCGT GGAGGTCCCC TTCCCTGCTCT CCAGCAAGTA

35 51 CGATGAGCCC AGCCGCCAGG TCATCCTGGA GGCTCTTGGC GAGTTTGAAC

101 GTTCCACGTG CATCAGGTTT GTCACCTATC AGGACCAGAG AGACTTCATT

151 TCCATCATCC CCATGTATGG GTGCTCTCG AGTGTGGGGC GCAGTGGAGG

40 201 GATGCAGGTG GTCTCCCTGG CGCCCACGTG TCTCCAGAAG GGCCGGGGCA

251 TTGTCCCTCA TGAGCTCATG CATGTGCTGG GCTTCTGGCA CGAGCACACG

45 301 CGGGCCGACC GGGACCGCTA TATCCGTGTC AACTGGAACG AGATCCTGCC

351 AGGCTTGAA ATCAACTTCA TCAAGTCTCA GAGCAGCAAC ATGCTGACGC

401 CCTATGACTA CTCCCTGTG ATGCACTATG GGAGGCTCGC CTTCAGCCGG

451 CGTGGGCTGC CCACCATCAC ACCACTTGG GCCCCCAGTG TCCACATCGG

50 501 CCAGCGATGG AACCTGAGTG CCTCGGACAT CACCCGGGTC CTCAAACCTCT

55 551 ACGGCTGCAG CCCAAGTGGC CCCAGGCCGT GTGGGAGAGG GTCCCATGCC

601 CACAGCACTG GTAGGAGCCC CGCCCCGGCC TCCCTATCTC TGCAGCGGCT

651 TTTGGAGGCA CTGTCGGCGG AATCCAGGAG CCCCCACCCC AGTGGTTCCA

60 701 GTGCAGGGAGG CCAGCCCCGT CCTGCAGGGC CTGGGGAGAG CCCACATGGG

751 TGGGAGTCCC CTGCCCTGAA AAAGCTCAGT GCAGAGGCCT CGGCAAGGCA  
801 GCCTCAGACC CTAGCTTCCT CCCCAAGATC AAGGCCTGGA GCAGGTGCC  
5 851 CCGGTGTTGC TCAGGAGCAG TCCTGGCTGG CCGGAGTGTC CACCAAGCCC  
901 ACAGTCCCCT CTTCAGAAGC AGGAATCCAG CCAGTCCCTG TCCAGGGAAAG  
951 CCCAGCTCTG CCAGGGGGCT GTGTACCTAG AAATCATTTC AAGGGGATGT  
10 1001 CCGAAGAT

**SEQ ID NO: 14 (INSP005A full protein sequence)**

15 1 MGGSGVVEVP FLLSSKYDEP SRQVILEALA EFERSTCIRF VTYQDQRDFI  
51 SIIPMYGCFS SVGRSGGMQV VSLAPTCQLQK GRGIVLHELM HVLGFWHEHT  
101 RADRDYIRV NWNEILPGFE INFIKSQSSN MLTPYDYSSV MHYGRLAFSR  
20 151 RGLPTITPLW APSVHIGQRW NLSASDITRV LKLYGCSPSG PRPRGRGSHA  
201 HSTGRSPAPA SLSLQRLLEA LSAESRSPDP SGSSAGGQPV PAGPGESEPHG  
25 251 WESPALKKLS AEASARQPQT LASSPRSRPG AGAPGVAQEQ SWLAGVSTKP  
301 TVPSSEAGIQ PVPVQGSPAL PGCCVPRNHF KGMSED

**SEQ ID NO: 15 (INSP005B nucleotide sequence exon 1)**

30 1 ATGGAGGGTG TAGGGGGTCT CTGGCCTTGG GTGCTGGGT TGCTCTCCTT  
51 GCCAG

**SEQ ID NO: 16 (INSP005B protein sequence exon 1)**

35 1 MEGVGGLWPW VLGLLSLPG

**SEQ ID NO: 17 (INSP005B nucleotide sequence exon 2)**

1 GTGTGATCCT AGGAGCGCCC CTGGCCTCCA GCTGCGCAGG AGCCTGTGGT  
51 ACCAGCTTCC CAGATGGCCT CACCCCTGAG GGAACCCAGG CCTCCGGGGA  
40 101 CAAGGACATT CCTGCAATTAA ACCAAG

**SEQ ID NO: 18 (INSP005B protein sequence exon 2)**

45 1 VILGAPLASS CAGACGTSFP DGLTPEGTQA SGDKDIPAIN QG

**SEQ ID NO: 19 (INSP005B nucleotide sequence exon 3)**

1 GGCTCATCCT GGAAGAAACC CCAGAGAGCA GCTTCCTCAT CGAGGGGGAC  
51 ATCATCCGGC CG  
50

**SEQ ID NO: 20 (INSP005B protein sequence exon 3)**

1 LILEETPESS FLIEGDIIRP

**SEQ ID NO: 21 (INSP005B nucleotide sequence exon 4)**

55 1 AGTCCCTTCC GACTGCTGTC AGCAACCAGC AACAAATGGC CCATGGGTGG

51 TAGTGGTGTC GTGGAGGTCC CCTTCCTGCT CTCCAGCAAG TACG

**SEQ ID NO: 22 (INSP005B protein sequence exon 4)**

5 1 SPFRLLSATS NKWPMGGSGV VEVPFLLSSK YD

**SEQ ID NO: 23 (INSP005B nucleotide sequence exon 5)**

1 ATGAGCCCAG CCGCCAGGTC ATCCTGGAGG CTCTTGCGGA GTTTGAACGT

10 51 TCCACGTGCA TCAGGTTTGT CACCTATCAG GACCAGAGAG ACTTCATTTC

101 CATCATCCCC ATGTATGG

**SEQ ID NO: 24 (INSP005B protein sequence exon 5)**

15 1 EPSRQVILEA LAEFERSTCI RFVTYQDQRD FISIIPMYG

**SEQ ID NO: 25 (INSP005B nucleotide sequence exon 6)**

1 GTGCTTCTCG AGTGTGGGGC GCAGTGGAGG GATGCAGGTG GTCTCCCTGG

20 51 CGCCCACGTG TCTCCAGAAG GGCCGGGGCA TTGTCCCTCA TGAGCTCATG

101 CATGTGCTGG GCTTCTGGCA CGAGCACACG CGGGCCGACC GGGACCGCTA

25 151 TATCCGTGTC AACTGGAACG AGATCCTGCC AG

**SEQ ID NO: 26 (INSP005B protein sequence exon 6)**

1 CFSSVGRSGG MQVVSLAPTC LQKGRGIVLH ELMHVLGFWH EHTRADRDY

30 51 IRVNWNEILP G

**SEQ ID NO: 27 (INSP005B nucleotide sequence exon 7)**

1 GCTTGAAAT CAACTTCATC AAGTCTCGGA GCAGCAACAT GCTGACGCC

35 51 TATGACTACT CCTCTGTGAT GCACTATGGG AG

**SEQ ID NO: 28 (INSP005B protein sequence exon 7)**

1 FEINFIKSRS SNMLTPYDYS SVMHYGR

**SEQ ID NO: 29 (INSP005B nucleotide sequence exon 8)**

40 1 GCTCGCCTTC AGCCGGCGTG GGCTGCCAC CATCACACCA CTTTGGGCC

51 CCAGTGTCCA CATCGGCCAG CGATGGAACC TGAGTGCCTC GGACATCACC

45 101 CGGGTCCTCA AACTCTACGG CTGCAGCCCA AGTGGCCCCA GGCCCCGTGG

151 GAGAG

**SEQ ID NO: 30 (INSP005B protein sequence exon 8)**

50 1 LAFSRRGLPT ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG

51 RG

**SEQ ID NO: 31 (INSP005B nucleotide sequence exon 9)**

1 GGTCCCATGC CCACAGCACT GGTAGGAGCC CCGCTCCGGC CTCCCTATCT  
 5 51 CTGCAGCGGC TTTGGAGGC ACTGTCGGCG GAATCCAGGA GCCCCGACCC  
 101 101 CAGTGGTTCC AGTGCAGGGAG GCCAGCCCCTGC TCCTGCAGGG CCTGGGGAGA  
 151 151 GCCCACATGG GTGGGAGTCC CCTGCCCTGA AAAAGCTCAG TGCAGAGGCC  
 201 10 201 TCGGCAAGGC AGCCTCAGAC CCTAGCTTCC TCCCCAAGAT CAAGGCCTGG  
 251 251 AGCAGGTGCC CCCGGTGTG CTCAGGAGCA GTCCCTGGCTG GCCGGAGTGT  
 301 15 301 CCACCAAGCC CACAGTCCC TCTTCAGAAG CAGGAATCCA GCCAGTCCT  
 351 351 GTCCAGGGAA GCCCAGCTCT GCCAGGGGGC TGTGTACCTA GAAATCATT  
 401 401 CAAGGGGATG TCCGAAGAT

**20 SEQ ID NO: 32 (INSP005B protein sequence exon 9)**

1 SHAHSTGRSP APASLSLQRL LEALSAESRS PDPSSGSSAGG QPVPAGPGES  
 5 51 PHGWESPALK KLSAEASARQ PQTLOSSPRS RPGAGAPGVA QEWSLAGVS  
 25 101 TKPTVPSSEA GIQPVPVQGS PALPGCVPR NHFKGMSED

**SEQ ID NO: 33 (INSP005B full nucleotide sequence)**

1 ATGGAGGGTG TAGGGGGTCT CTGGCCTTGG GTGCTGGGTG TGCTCTCCTT  
 30 51 GCCAGGTGTG ATCCTAGGAG CGCCCCCTGGC CTCCAGCTGC GCAGGAGCCT  
 101 101 GTGGTACCAAG CTTCCCAGAT GGCCTCACCC CTGAGGGAAC CCAGGCCTCC  
 151 151 GGGGACAAGG ACATTCCCTGC AATTAACCAA GGGCTCATCC TGGAAGAAC  
 201 201 CCCAGAGAGC AGCTTCCCTCA TCGAGGGGGC CATCATCCGG CCGAGTCCT  
 251 251 TCCGACTGCT GTCAGCAACC AGCAACAAAT GGCCCATGGG TGGTAGTGGT  
 301 40 301 GTCGTGGAGG TCCCCTTCCT GCTCTCCAGC AAGTACGATG AGCCCAGCCG  
 351 351 CCAGGTCATC CTGGAGGCTC TTGCGGAGTT TGAACGTTCC ACGTGCATCA  
 401 401 GGTTTGTAC CTTACAGGAC CAGAGAGACT TCATTTCAT CATCCCCATG  
 451 451 TATGGGTGCT TCTCGAGTGT GGGGCGCAGT GGAGGGATGC AGGTGGTCTC  
 501 501 CCTGGCGCCC ACGTGTCTCC AGAAGGGCCG GGGCATTGTC CTTCATGAGC  
 551 551 TCATGCATGT GCTGGGCTTC TGGCACGAGC ACACGGGGGC CGACCGGGAC  
 601 601 CGCTATATCC GTGTCAACTG GAACGAGATC CTGCCAGGCT TTGAAATCAA  
 651 651 CTTCATCAAG TCTCGGAGCA GCAACATGCT GACGCCCTAT GACTACTCCT  
 701 701 CTGTGATGCA CTATGGGAGG CTCGCCTTCA GCCGGCGTGG GCTGCCACC  
 751 751 ATCACACCAAC TTTGGGCCCT CAGTGTCCAC ATCGGCCAGC GATGGAACCT  
 801 801 GAGTGCCTCG GACATCACCC GGGTCCTCAA ACTCTACGGC TGCAGCCCAA  
 851 851 GTGGCCCCAG GCCCGCGTGGG AGAGGGTCCC ATGCCACAG CACTGGTAGG  
 901 901 AGCCCCGCTC CGGCCTCCCT ATCTCTGCAG CGGCTTTGG AGGCACTGTC

951 GGCGGAATCC AGGAGCCCCG ACCCCAGTGG TTCCAGTGC GGGAGGCCAGC  
 1001 CCGTTCTGC AGGGCCTGGG GAGAGCCCAC ATGGGTGGGA GTCCCCTGCC  
 5 1051 CTGAAAAAGC TCAGTGCAGA GGCCTCGGCA AGGCAGCCTC AGACCCCTAGC  
 1101 TTCCTCCCCA AGATCAAGGC CTGGAGCAGG TGCCCCCGGT GTTGCTCAGG  
 10 1151 AGCAGTCCTG GCTGGCCGGA GTGTCCACCA AGCCCACAGT CCCATCTTCA  
 1201 GAAGCAGGAA TCCAGCCAGT CCCTGTCCAG GGAAGCCCAG CTCTGCCAGG  
 15 1251 GGGCTGTGTA CCTAGAAATC ATTTCAAGGG GATGTCCGAA GAT

**SEQ ID NO: 34 (INSP005B full protein sequence)**

1 MEGVGGLWPW VLGLLSLPGV ILGAPLASSC AGACGTSFPD GLTPEGTQAS  
 5 20 51 GDKDIPAINQ GLILEETPES SFLIEGDIIR PSPFRLLSAT SNKWPMSGSG  
 101 VVEVPFLLSS KYDEPSRQVI LEALAEFERS TCIRFVTYQD QRDFISIIPM  
 151 YGCFSSVGRS GGMQVVLAP TCLQKRGIV LHELMHVLGF WHEHTRADRD  
 25 201 RYIRVNNEI LPGFEINFIK SRSSNMLTPY DYSSVMHYGR LAFSRRGLPT  
 251 ITPLWAPS VH IQRWNLSAS DITRVLKLYG CSPSGPRPRG RGSHAHSTGR  
 30 301 SPAPASLSLQ RLLEALSAES RSPDPSGSSA GGQPVPAGPG ESPHWESPA  
 351 LKKLSAEASA RQPQTASSP RSRPGAGAPG VAQEWSLWAG VSTKPTVPSS  
 401 EAGIQPVVQ GSPALPGGCV PRNHFKGMS D

**35 SEQ ID NO: 35 (INSP005b mature nucleotide sequence)**

1 GCGCCCTGG CCTCCAGCTG CGCAGGAGCC TGTGGTACCA GCTTCCCAGA  
 5 51 TGGCCTCACC CCTGAGGGAA CCCAGGCCTC CGGGGACAAG GACATTCCCTG  
 101 CAATTAACCA AGGGCTCATC CTGGAAGAAA CCCCAGAGAG CAGCTTCCTC  
 151 ATCGAGGGGG ACATCATCCG GCCGAGTCCC TTCCGACTGC TGTCAAGAAC  
 40 201 CAGCAACAAA TGGCCCATGG GTGGTAGTGG TGTCGTGGAG GTCCCCTTCC  
 251 TGCTCTCCAG CAAGTACGAT GAGCCCAGCC GCCAGGTCA CCTGGAGGCT  
 301 CTTGCGGAGT TTGAACGTT CACGTGCATC AGGTTTGTCA CCTATCAGGA  
 351 CCAGAGAGAC TTCATTTCA TCATCCCCAT GTATGGGTGC TTCTCGAGTG  
 401 TGGGGCGCAG TGGAGGGATG CAGGTGGTCT CCCTGGCGCC CACGTGTCTC  
 45 451 CAGAACGGCC GGGGCATTGT CCTTCATGAG CTCATGCATG TGCTGGCTT  
 501 CTGGCACGAG CACACGGGG CCGACCGGGA CCGCTATATC CGTGTCAACT  
 551 GGAACGAGAT CCTGCCAGGC TTTGAAATCA ACTTCATCAA GTCTCGGAGC  
 601 AGCAACATGC TGACGCCCTA TGACTACTCC TCTGTGATGC ACTATGGGAG  
 651 GCTCGCCTTC AGCCGGCGTG GGCTGCCAC CATCACACCA CTTTGGGCC  
 50 701 CCAGTGTCCA CATCGGCCAG CGATGGAACC TGAGTGCCTC GGACATCACC  
 751 CGGGTCTCA AACTCTACGG CTGCAGCCCA AGTGGCCCCA GGCCCCGTGG  
 801 GAGAGGGTCC CATGCCACCA GCACTGGTAG GAGCCCCGCT CCGGCCTCCC  
 851 TATCTCTGCA GCGGTTTG GAGGCAGTGT CGGCGGAATC CAGGAGCCCC

901 GACCCCAGTG GTTCCAGTGC GGGAGGCCAG CCCGTTCTTG CAGGGCCTGG  
 951 GGAGAGCCCA CATGGGTGGG AGTCCCCTGC CCTGAAAAAG CTCAGTGCAG  
 1001 AGGCCTCGGC AAGGCAGCCT CAGACCCCTAG CTTCCCTCCCC AAGATCAAGG  
 1051 CCTGGAGCAG GTGCCCGGG TGTTGCTCAG GAGCAGTCCT GGCTGGCCGG  
 5 1101 AGTGTCCACC AAGCCCACAG TCCCCATCTTC AGAACAGGA ATCCAGGCCAG  
 1151 TCCCTGTCCA GGGAAAGCCCA GCTCTGCCAG GGGGCTGTGT ACCTAGAAAT  
 1201 CATTCAAGG GGATGTCCGA AGAT

**SEQ ID NO: 36 (INSP005b mature polypeptide sequence)**

10 1 APLASSCAGA CGTSFPDGLT PEGTQASGDK DIPAINQGLI LEETPESSFL  
 51 IEGDIIRPSP FRLLSATSNK WPMGGSGVVE VPFLSSKYD EPSRQVILEA  
 101 LAEFERSTCI RFVTYQDQRD FISIIPMYGC FSSVGRSGGM QVVSLLAPTCL  
 151 QKGRGIVLHE LMHVLGFHWE HTRADRDYI RVNWNEILPG FEINFIKSRS  
 201 SNMLTPYDYS SVMHYGRLAF SRRGLPTITP LWAPSVHIGQ RWNLASDIT  
 15 251 RVLKLYGCSP SGPRPRGRGS HAHSTGRSPA PASLSLQRLL EALSAESRSP  
 301 DPSGSSAGGQ PVPAGPGESP HGWESPALKK LSAEASARQP QTLASSPRSR  
 351 PGAGAPGVAQ EQSWLAGVST KPTVPSSEAG IQPVPVQGSP ALPGGCVPRN  
 401 HFKGMSED

**20 SEQ ID NO: 37 (INSP005 Predicted Polypeptide Sequence)**

1 1 MLRLWDFNPG GALSDLALGL RGMEEGGYSC AGACGTSFPD GLTPEGTQAS GDKDIPAINQ  
 61 61 GLILEEETPES SFLIEGDIIR PSPFRLLSAT SNKWPMSGSG VVEVPFLSS KYDEPSHQVI  
 121 121 LEALAEFERS TCIRFVTYQD QRDFISIIPM YGCFSSVGRS GGMQVVSLLAP TCLQKGRGIV  
 181 181 LHELMHVLGF WHEHTRADRD RYIRVNWNEI LPGFEINFIK SQSSNMLTPY DYSSVMHYGR  
 25 241 241 LAFSRRGLPT ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG RGEWHGRKV

**SEQ ID NO: 38 (pCR4 TOPO IPAAA78836-1 plasmid nucleotide sequence)**

1 1 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC  
 61 61 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC  
 30 121 121 TCACTCATTAA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA  
 181 181 TTGTGAGCGG ATAACAATTTC CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTCA  
 241 241 GAATTAAACCC TCACTAAAGG GACTAGTCCT GCAGGTTAAACCGAATTTCGC CCTTAGCCAC  
 301 301 AGGCTTAATC TTGGACATC CCCTTGAAAT GATTTCTAGG TACACAGCCC CCTGGCAGAG  
 361 361 CTGGGCTTCC CTGGACAGGG ACTGGCTGGA TTCCCTGCTTC TGAAGATGGG ACTGTGGGCT  
 35 421 421 TGGTGGACAC TCCGGCCAGC CAGGACTGCT CCTGAGCAAC ACCGGGGGCA CCTGCTCCAG  
 481 481 GCCTTGATCT TGGGGAGGAA GCTAGGGTCT GAGGCTGCCT TGCCGAGGCC TCTGCACTGA  
 541 541 GCTTTTCAG GGCAGGGGAC TCCCACCCAT GTGGGCTCTC CCCAGGCCCT GCAGGAACGG  
 601 601 GCTGGCCTCC CGCACTGGAA CCACTGGGTT CGGGGCTCCT GGATTCCGCC GACAGTGCCT  
 661 661 CCAAAAGCCG CTGCAGAGAT AGGGAGGCCG GGGCGGGGCT CCTACCAGTG CTGTGGCCT

721 GGGACCCTCT CCCACGGGGC CTGGGGCAC TTGGGCTGCA GCCGTAGAGT TTGAGGACCC  
 781 GGGTGATGTC CGAGGCACTC AGGTTCCATC GCTGGCCGAT GTGGACACTG GGGGCCAAA  
 841 GTGGTGTGAT GGTGGGCAGC CCACGCCGGC TGAAGGGAG CCTCCCATAG TGCATCACAG  
 901 AGGAGTAGTC ATAGGGCGTC AGCATGTTGC TGCTCTGAGA CTTGATGAAG TTGATTTCAA  
 5 961 AGCCTGGCAG GATCTCGTTC CAGTTGACAC GGATATAGCG GTCCCGGTG GCCCACGTGT  
 1021 GCTCGTGCCTA GAAGCCCAGC ACATGCATGA GCTCATGAAG GACAATGCC CCGCCCTTCT  
 1081 GGAGACACGT GGGGCCAGG GAGACCACCT GCATCCCTCC ACTGCGCCCC ACACTCGAGA  
 1141 AGCACCCATA CATGGGGATG ATGGAAATGA AGTCTCTCTG GTCCCTGATAG GTGACAAACC  
 1201 TGATGCACGT GGAACGTTCA AACTCCGCAA GAGCCTCCAG GATGACCTGG CGGCTGGCT  
 10 1261 CATCGTACTT GCTGGAGAGC AGGAAGGGGA CCTCCACGAC ACCACTACCA CCCATGGGCC  
 1321 ATTTGTTGCT GGTTGCTGAC AGAAGGGCGA ATTTCGCGGCC GCTAAATTCA ATTTCGCCCTA  
 1381 TAGTGAGTCG TATTACAATT CACTGGCCGT CGTTTACAA CGTCGTGACT GGGAAAACCC  
 1441 TGGCGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT GGCGTAATAG  
 1501 CGAAGAGGCC CGCACCGATC GCCCCCTCCCA ACAGTTGCAGC AGCCTATACG TACGGCAGTT  
 15 1561 TAAGGTTTAC ACCTATAAAA GAGAGAGCCG TTATCGTCTG TTTGTGGATG TACAGAGTGA  
 1621 TATTATTGAC ACGCCGGGGC GACGGATGGT GATCCCCCTG GCCAGTGCAC GTCTGCTGTC  
 1681 AGATAAAAGTC TCCCGTGAAC TTTACCCGGT GGTGCATATC GGGGATGAAA GCTGGCGCAT  
 1741 GATGACCACC GATATGGCCA GTGTGCCGGT CTCCGTTATC GGGGAAGAAG TGGCTGATCT  
 1801 CAGCCACCGC GAAAATGACA TCAAAAACGC CATTACCTG ATGTTCTGGG GAATATAAAT  
 20 1861 GTCAGGCATG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTCACGT AGAAAGCCAG  
 1921 TCCGCAGAAA CCGTGCTGAC CCCGGATGAA TGTCAGCTAC TGGGCTATCT GGACAAGGG  
 1981 AAACGCAAGC GCAAAGAGAA AGCAGGTAGC TTGCAGTGGG CTTACATGGC GATAGCTAGA  
 2041 CTGGGCAGTT TTATGGACAG CAAGCGAACC GGAATTGCCA GCTGGGGCGC CCTCTGGTAA  
 2101 GGTTGGGAAG CCCTGCAAAG TAAACTGGAT GGCTTTCTCG CCGCCAAGGA TCTGATGGCG  
 25 2161 CAGGGGATCA AGCTCTGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA  
 2221 TGGATTGAC GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGC  
 2281 ACAACAGACA ATCGGCTGCT CTGATGCCGC CGTGTCCGG CTGTCAGCGC AGGGCGCCC  
 2341 GGTTCTTTT GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAAG ACGAGGCAGC  
 2401 GCGGCTATCG TGGCTGGCCA CGACGGCGT TCCTTGCAGC GCTGTGCTCG ACGTTGTCAC  
 30 2461 TGAAGCGGGGA AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC  
 2521 TCACCTTGCT CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCC GGCCTGCATAC  
 2581 GCTTGATCCG GCTACCTGCC CATTGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG  
 2641 TACTCGGATG GAAGCGGGTC TTGTCATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT  
 2701 CGCGCCAGCC GAACTGTTCG CCAGGCTCAA GGCGAGCATG CCCGACGGCG AGGATCTCGT  
 35 2761 CGTGACCCAT GGCGATGCC GCTTGCCGAA TATCATGGT GAAAATGCC GCTTTCTGG  
 2821 ATTCACTCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC  
 2881 CCGTGATATT GCTGAAGAGC TTGGCGCGA ATGGGCTGAC CGCTTCCCTCG TGCTTTACGG  
 2941 TATCGCCGCT CCCGATTCCG AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTCTG  
 3001 AATTATTAAC GCTTACAATT TCCTGATGCC GTATTTCTC CTTACGCATC TGTGCAGGTAT  
 40 3061 TTCACACCGC ATACAGGTGG CACTTTCCG GGAAATGTGC GCGGAACCCC TATTGTTA  
 3121 TTTTTCTAAA TACATTCAA TATGTATCCG CTCATGAGAC AATAACCCCTG ATAAATGCTT

3181 CAATAATATT GAAAAAGGAA GAGTATGAGT ATTCAACATT TCCGTGTCGC CCTTATTC  
 3241 TTTTTGCGG CATTTCGCCT TCCTGTTTT GCTCACCCAG AAACGCTGGT GAAAGTAAA  
 3301 GATGCTGAAG ATCAGTTGGG TGCACGAGTG GGTTACATCG AACTGGATCT CAACAGCGGT  
 3361 AAGATCCTTG AGAGTTTCG CCCCCGAAGAA CGTTTCCAA TGATGAGCAC TTTAAAGTT  
 5 3421 CTGCTATGTG GCGCGGTATT ATCCCGTATT GACGCCGGC AAGAGCAACT CGGTCGCCGC  
 3481 ATACACTATT CTCAGAATGA CTTGGTTGAG TACTCACCAAG TCACAGAAAA GCATCTTACG  
 3541 GATGGCATGA CAGTAAGAGA ATTATGCAGT GCTGCCATAA CCATGAGTGA TAACACTGCG  
 3601 GCCAACTTAC TTCTGACAAC GATCGGAGGA CCGAAGGAGC TAACCGCTTT TTTGCACAAC  
 3661 ATGGGGGATC ATGTAACTCG CCTTGATCGT TGGGAACCGG AGCTGAATGA AGCCATACCA  
 10 3721 AACGACGAGC GTGACACCAC GATGCCGTGA GCAATGGCAA CAACGTTGCG CAAACTATTA  
 3781 ACTGGCGAAC TACTTACTCT AGCTTCCCGG CAACAATTAA TAGACTGGAT GGAGGCGGAT  
 3841 AAAGTTGCAG GACCACTTCT GCGCTCGGCC CTTCCGGCTG GCTGGTTTAT TGCTGATAAA  
 3901 TCTGGAGCCG GTGAGCGTGG GTCTCGCGGT ATCATTGCAG CACTGGGGCC AGATGGTAAG  
 3961 CCCTCCCGTA TCGTAGTTAT CTACACGACG GGGAGTCAGG CAACTATGGA TGAACGAAAT  
 15 4021 AGACAGATCG CTGAGATAGG TGCCCTCACTG ATTAAGCATT GGTAACTGTC AGACCAAGTT  
 4081 TACTCATATA TACTTTAGAT TGATTTAAA CTTCATTTTT AATTAAAAG GATCTAGGTG  
 4141 AAGATCCTTT TTGATAATCT CATGACCAAA ATCCCTAAC GTGAGTTTC GTTCCACTGA  
 4201 GCGTCAGACC CCGTAGAAAA GATCAAAGGA TCTTCTTGAG ATCCTTTTT TCTGCGCGTA  
 4261 ATCTGCTGCT TGCAAACAAA AAAACCACCG CTACCAGCGG TGGTTTGTGTT GCCGGATCAA  
 20 4321 GAGCTACCAA CTCTTTTCC GAAGGTAAC GGCTTCAGCA GAGCGCAGAT ACCAAATACT  
 4381 GTCCTCTAG TGTAGCCGTA GTTAGGCCAC CACTTCAAGA ACTCTGTAGC ACCGCCTACA  
 4441 TACCTCGCTC TGCTAATCCT GTTACCAAGTG GCTGCTGCCA GTGGCGATAA GTCGTGTCTT  
 4501 ACCGGGTTGG ACTCAAGACG ATAGTTACCG GATAAGGCAG AGCGGTCGGG CTGAACGGGG  
 4561 GGTTCGTGCA CACAGCCCAG CTTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG  
 25 4621 CGTGAGCTAT GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGGGGACAG GTATCCGGTA  
 4681 AGCGGCCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC CAGGGGGAAA CGCCTGGTAT  
 4741 CTTTATAGTC CTGTCGGGTT TCGCCACCTC TGACTTGAGC GTCGATTTTT GTGATGCTCG  
 4801 TCAGGGGGGC GGAGCCTATG GAAAAACGCC AGCAACGCCGG CCTTTTTACG GTTCCTGGGC  
 4861 TTTTGCTGGC CTTTGCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC  
 30 4921 CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC CGAGCGCAGC  
 4981 GAGTCAGTGA GCGAGGAAGC GGAAG

**SEQ ID NO: 39 (XpCR4TOPO IPAAA78836-2 plasmid nucleotide sequence)**

1 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC  
 35 61 ACGACAGGTT TCCCAGCTGG AAAGCGGGCA GTGAGGCCAA CGCAATTAAAT GTGAGTTAGC  
 121 TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA  
 181 TTGTGAGCGG ATAACAATT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTCA  
 241 GAATTAAACCC TCACTAAAGG GACTAGTCCT GCAGGTTAA ACGAATTCCGC CCTTAGCCAC  
 301 AGGCTTAATC TTCCGGACATC CCCTTGAAAT GATTCTAGG TACACAGCCC CCTGGCAGAG  
 40 361 CTGGGCTTCC CTGGACAGGG ACTGGCTGGA TTCCCTGCTTC TGAAGATGGG ACTGTGGGCT

421 TGGTGGACAC TCCGGCCAGC CAGGACTGCT CCTGAGCAAC ACCGGGGCA CCTGCTCCAG  
 481 GCCTTGATCT TGGGGAGGAA GCTAGGGTCT GAGGCTGCC TGCCGAGGCC TCTGCACTGA  
 541 GCTTTTTCAG GGCAGGGGAC TCCCACCCAT GTGGGCTCTC CCCAGGCCCT GCAGGAACGG  
 601 GCTGGCCTCC CGCACTGGAA CCACTGGGT CGGGGCTCCT GGATTCCGCC GACAGTGCCT  
 5 661 CCAAAAGCCG CTGCAGAGAT AGGGAGGCCG GAGCGGGGCT CCTACCACTG CTGTGGCAT  
 721 GGGACCCCTCT CCCACGGGC CTGGGCCAC TTGGGCTGCA GCCGTAGAGT TTGAGGACCC  
 781 GGGTGATGTC CGAGGCACTC AGGTTCCATC GCTGGCCGAT GTGGACACTG GGGGCCAAA  
 841 GTGGTGTGAT GGTGGGCAGC CCACGCCGGC TGAAGGCGAG CCTCCCATAG TGCATCACAG  
 901 AGGAGTAGTC ATAGGGCTC AGCATGTTGC TGCTCCGAGA CTTGATGAAG TTGATTCA  
 10 961 AGCCTGGCAG GATCTCGTTC CAGTTGACAC GGATATAGCG GTCCCGGTG GCCC CGGTGTT  
 1021 GCTCGTGCCA GAAGCCCAGC ACATGCATGA GCTCATGAAG GACAATGCC GGGCCCTTCT  
 1081 GGAGACACGT GGGCGCCAGG GAGACCACCT GCATCCCTCC ACTGCGCCCC ACAC TCGAGA  
 1141 AGCACCCATA CATGGGGATG ATGAAATGA AGTCTCTCTG GTCCCTGATAG GTGACAAACC  
 1201 TGATGCACGT GGAACGTTCA AACTCCGAA GAGCCTCCAG GATGACCTGG CGGCTGGCT  
 15 1261 CATCGTACTT GCTGGAGAGC AGGAAGGGGA CCTCCACGAC ACCACTACCA CCCATGGCC  
 1321 ATTTGTTGCT GGTTGCTGAC AGCAGTCGGA AGGGACTCGG CCGGATGATG TCCCCCTCGA  
 1381 TGAGGAAGCT GCTCTCTGGG GTTTCTTCCA GGATGAGGCC TTGGTTAATT GCAGGAATGT  
 1441 CCTTGTCCTCC GGAGGCCTGG GTTCCCTCAG GGGTGAGGCC ATCTGGGAAG CTGGTACAC  
 1501 AGGCTCCTGC GCAGCTGGAG GCCAGGGCG CTCCTAGGAT CACACCTGGC AAGGAGAGCA  
 20 1561 GACCCAGCAC CCAAGGCCAG AGACCCCTA CACCCCTCCAT GGTAGAAAGG GCGAATTGCG  
 1621 GGCCGCTAAA TTCAATTGCG CCTATAGTGA GTCGTATTAC AATTCACTGG CCGTCGTTT  
 1681 ACAACGTCGT GACTGGAAA ACCCTGGCGT TACCCAACCTT AATCGCCTTG CAGCACATCC  
 1741 CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACC GATGCCCTT CCCAACAGTT  
 1801 GCGCAGCCTA TACGTACGGC AGTTAAGGT TTACACCTAT AAAAGAGAGA GCCGTTATCG  
 25 1861 TCTGTTGTC GATGTACAGA GTGATATTAT TGACACGCCG GGGCGACGGA TGGTGATCCC  
 1921 CCTGGCCAGT GCACGTCTGC TGTCAGATAA AGTCTCCGT GAAC TTTACC CGGTGGTGCA  
 1981 TATCGGGGAT GAAAGCTGGC GCATGATGAC CACCGATATG GCCAGTGTGC CGGTCTCCGT  
 2041 TATCGGGGAA GAAGTGGCTG ATCTCAGCCA CCGCGAAAAT GACATCAAAA ACGCCATTAA  
 2101 CCTGATGTTG TGGGAATAT AAATGTCAGG CATGAGATTA TCAAAAAGGA TCTTCACCTA  
 30 2161 GATCCTTTTC ACGTAGAAAG CCAGTCGC AAACGGTGC TGACCCCGGA TGAATGTCAG  
 2221 CTACTGGCT ATCTGGACAA GGGAAAACGC AAGCGCAAAG AGAAAGCAGG TAGCTTGCA  
 2281 TGGGCTTACA TGGCGATAGC TAGACTGGC GTTATGG ACAGCAAGCG AACCGGAATT  
 2341 GCCAGCTGGG CGCCCTCTG GTAAAGGTGG GAAGCCCTGC AAAGTAAACT GGATGGCTTT  
 2401 CTCGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGCTCT GATCAAGAGA CAGGATGAGG  
 35 2461 ATCGTTTCGC ATGATTGAAC AAGATGGATT GCACGCAGGT TCTCCGGCCG CTTGGGTGGA  
 2521 GAGGCTATTG GCCTATGACT GGGCACAAACA GACAATCGGC TGCTCTGATG CCGCCGTGTT  
 2581 CCGGCTGTCA GCGCAGGGC GCCCGGTTCT TTTGTCAAG ACCGACCTGT CCGGTGCCCT  
 2641 GAATGAACTG CAAGACGAGG CAGCGCGGCT ATCGTGGCTG GCCACGACGG GCGTTCTTG  
 2701 CGCAGCTGTG CTCGACGTTG TCACTGAAGC GGGAAAGGGAC TGGCTGCTAT TGGCGAAGT  
 40 2761 GCGGGGGCAG GATCTCCTGT CATCTCACCT TGCTCCTGCC GAGAAAGTAT CCATCATGGC  
 2821 TGATGCAATG CGGGCGCTGC ATACGCTTGA TCCGGCTACC TGCCCATTCG ACCACCAAGC

2881 GAAACATCGC ATCGAGCGAG CACGTACTCG GATGGAAGCC GGTCTTGTG ATCAGGATGA  
 2941 TCTGGACGAA GAGCAGTCAGG GGCTCGCGCC AGCCGAAC TG TTGCGCAGGC TCAAGGCGAG  
 3001 CATGCCGAC GGCAGGGATC TCGTCGTGAC CCATGGCGAT GCCTGCTTGC CGAATATCAT  
 3061 GGTGGAAAAT GGCGCTTT CTGGATTCA CGACTGTGGC CGGCTGGTG TGGCGGACCG  
 5 3121 CTATCAGGAC ATAGCGTTGG CTACCCGTGA TATTGCTGAA GAGCTTGGCG GCGAATGGC  
 3181 TGACCGCTTC CTCGTGCTTT ACGGTATCGC CGCTCCGAT TCGCAGCGA TCGCCTCTA  
 3241 TCGCCTCTT GACGAGTTCT TCTGAATTAT TAACGCTTAC AATTTCCTGA TGCAGTATTT  
 3301 TCTCCTTACG CATCTGTGCG GTATTCACA CCGCATAACAG GTGGCACTTT TCGGGGAAAT  
 3361 GTGCGCGGAA CCCCTATTIG TTTATTTTC TAAATACATT CAAATATGTA TCCGCTCATG  
 10 3421 AGACAAATAAC CCTGATAAAAT GCTTCATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA  
 3481 CATTTCGGT TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC  
 3541 CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC  
 3601 ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTGCCCCGA AGAACGTTT  
 3661 CCAATGATGA GCACTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC  
 15 3721 GGGCAAGAGC AACTCGGTG CCGCATAACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA  
 3781 CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC  
 3841 ATAACCATGA GTGATAACAC TCGGCCAAC TTACTCTGA CAACGATCGG AGGACCGAAG  
 3901 GAGCTAACCG CTTTTTGCA CAACATGGGG GATCATGTAA CTGGCCTTGA TCGTTGGGAA  
 3961 CCGGAGCTGA ATGAAGCCAT ACCAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG  
 20 4021 GCAACAAACGT TCGCCTAACT ATTAACCTGGC GAAACTCTTA CTCTAGCTTC CCGGCAACAA  
 4081 TTAATAGACT GGATGGAGGC GGATAAAAGTT GCAGGACAC TTCTGCGCTC GGCCCTTCCG  
 4141 GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT  
 4201 GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT  
 4261 CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG  
 25 4321 CATTGGTAAC TGTCAAGACCA AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT  
 4381 TTTTAATTAA AAAGGATCTA GGTGAAGATC CTTTTGATA ATCTCATGAC CAAAATCCCT  
 4441 TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT  
 4501 TGAGATCCTT TTTTCTGCG CGTAATCTGC TGCTTGAAA CAAAAAAACC ACCGCTACCA  
 4561 GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC  
 30 4621 AGCAGAGCGC AGATACAAA TACTGTCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC  
 4681 AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT  
 4741 GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG  
 4801 GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC  
 4861 TACACCGAAC TGAGATACCT ACAGCGTGA CGTATGAGAAA GCGCCACGCT TCCCAGAGGG  
 35 4921 AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCCGAA CAGGAGAGCG CACGAGGGAG  
 4981 CTTCCAGGGG GAAACGCTG GTATCTTAT AGTCCTGTCG GGTTCGCCA CCTCTGACTT  
 5041 GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC  
 5101 GCGGCCTTTT TACGGTTCTT GGGCTTTGC TGGCCTTTG CTCACATGTT CTTTCCTGCG  
 5161 TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTG AGTGAAGCTGA TACCGCTCGC  
 40 5221 CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAG